Types and Some Causes of Nonsampling Errors in Household Surveys in Africa

Ben Kiregyera

1. Introduction

In many African countries, household surveys were initiated in the 1950s. The number of household surveys increased significantly in the 1960s following the attainment of independence. In addition to being ad hoc, uncoordinated and unintegrated, these surveys were subject-specific and limited in geographical coverage. The main subjects covered were income, expenditure, consumption, and demographic characteristics of households.

In the last decade or so, household surveys have come to play a central and strategic role in the development of national statistical programmes in Africa and other developing countries. Indeed, we have witnessed an “unprecedented expansion of household survey work in developing countries in response to rapidly expanding demands for current and detailed socio-economic data” (UNECA (1984)). Both subject matter and geographic coverage have been a part of this expansion. Currently a number of African countries are implementing national multipurpose household surveys on: income, consumption and expenditure, demographic aspects, agriculture, labour force, health and nutrition, handicraft activities, and household enterprise.

Eleven of these countries have continuing surveys as a part of the African Household Survey Capability Programme (AHSCP) which was launched in 1978 by the United Nations Economic Commission for Africa (UNECA).

In terms of data availability and accuracy, household surveys are the weakest element of African statistics. Household surveys are affected by both sampling and nonsampling errors. The former arise because a sample rather than an entire population is observed. Sampling errors tend not to be large for large-scale sample surveys and do not present great problems because many African household surveys are now performed on a large scale. Moreover, these surveys are often designed to ensure that sampling errors are brought under control. In household surveys, on the other hand, nonsampling errors constitute a great problem. These are systematic errors which are caused by a host of factors. Nonsampling errors are harder to control and measure and often contribute more to the total error than sampling errors.

This discussion is limited to the causes of some nonsampling errors at the data collection stage and the measures that are being taken to control these errors in African household survey programmes. The more widely reported and serious types of nonsampling errors that occur during data collection include: coverage errors, response errors, measurement errors, and interviewer bias.

1 Institute of Statistics and Applied Economics, Makerere University, P.O. Box 7062, Kampala, Uganda.
2. Coverage Errors

The main cause of coverage errors is defects in the operational sampling frame. These defects include inaccuracy, incompleteness, and outdatedness. These defects have many causes, and the more important causes are discussed below.

2.1. Population based sampling frames

Ready and usable national sampling frames of enumeration units, e.g., households, holdings, or dwellings are invariably nonexistent in Africa and the cost of their construction initio can be prohibitive. This makes sampling frame construction perhaps the most contentious and difficult task in household surveys. Many countries circumvent this problem by using population based sampling frames. Multi-stage cluster sampling designs based on population censuses are used. The census enumeration areas (EAs) function as primary sampling units (PSUs) and households or farm households (for agricultural surveys) function as ultimate sampling units (USUs). Such frames, however, become quickly outdated and do not form a good basis for preparing a sampling frame. It ought to be mentioned that in population censuses, all aspects of the field work, i.e., coverage, mapping, etc., have generally been of poor quality. The use of population based sampling frames leads to coverage errors. For instance, it has been estimated that a 10–15 % undercount is typical when EAs from population censuses are used as sampling frames for agricultural sample censuses and surveys.

2.2. Area based sampling frames

Some countries in Africa have resorted to using area-based sampling frames for their master samples as opposed to the population based frames discussed above. In these countries, cluster units have mainly consisted of villages and localities. These particular units are generally small and scattered over rural areas. This has created identification problems as well as mapping and field operation problems (Kiregyera (1987)). In some countries (e.g., Tanzania) where the village is used as the ultimate master sample unit, not all village boundaries are clear (Olsson (1984)). In many other countries, there have been problems of identifying villages and localities due to rapid changes in rural settlement patterns (fragmentation, disappearance, resettlement, regrouping). Further problems are caused by a lack of standard names for these localities.

In Gambia, for instance, locations are constantly changing and even a whole or a part of a village can move. This leads to a multiplicity of names for the same village. In Sierra Leone, field workers used their own initiative in spelling names of localities because no standard spellings existed (UNECA (1984)). And in Swaziland, the names the Swazi give to areas of land do not encompass readily definable boundaries and frequently one area is known by a number of names. It has also been reported that variations in spelling created special problems in frame construction for the 1984 household budget survey in Somalia (Olsson (1987)). The unavailability of adequate base maps generates mapping problems and the physical environment creates dispersion and discontinuous patterns in settlements. These factors combined with a shortage of cartographic skills at many African Central Statistical Offices (CSOs) have made frame construction a rigorous task indeed.

2.3. Definitional problems

Household surveys are planned and executed using internationally defined concepts some of which only roughly fit the actual conditions in Africa. Often, attempts are made to adjust the definitions of these concepts to the conditions in each country. It has, however, been observed that guidelines for such adjustments
are sometimes disregarded or misunderstood by enumerators. The "household", for instance, is used as a reporting unit in almost every household survey in Africa. This concept was defined by the United Nations Economic Commission for Africa (UNECA (1984)) in the hope to make the original United Nations (1980) definition of household cover all types of African households accommodating for the complex living and eating arrangements among husbands, wives, children, and relatives. This definition is as follows:

"A household is defined as a person or a group of persons who live together and eat together."

Attempts to adapt this definition to suit local conditions have had varying degrees of success (UNECA (1981a)). As a result of these adoptions, the definition of household varies from country to country according to local social structure, culture, and customs. It has been reported that despite the adjustments, it has been difficult to apply the definitions for the following reasons (UNECA (1981b)).

a. In Africa the notion of the family can be applied to a number of diverse family structures. The African definition of the family is broader and goes beyond the western concept of household and dwelling unit. A family can consist of natural relatives with varying degrees of relationship regardless of their residences.

b. In some societies the descendants of a single ancestor tend to live together or on the same compound. Thus, one family forms what would in the west be seen as separate households connected through the head of the family.

c. There is also polygamy, where one man can be the head of several households living in different residences. In some polygamous unions, each woman occupies a different residence with her children. Strict application of the concept of household in such cases is difficult. From a social point of view, the polygamous husband, his wives and children all form a single household.

The application of the concept of household has therefore been problematic in African household surveys and has led to coverage errors. For instance, when the definition has been used in agricultural surveys, it has led to erroneous results. In Mali, an entire family owns its lands jointly under the responsibility of a patriarch. Attempts to divide this type of farm into its different households has led to erroneous results. Since 1974 the household has been used as the ultimate sampling unit for all surveys (agricultural and nonagricultural) in Kenya and the suitability of using the household as a survey unit in agricultural surveys is now being questioned. In particular, it has been observed that use of the household tends to omit some holdings and leads to underestimation of acreage and production.

There have been attempts to adapt the African notion of a "holding" to statistical purposes, but this has also proved problematic. In the 1963–65 Ugandan agricultural sample census, the definition of holding seemed to have led to the erroneous enumeration of large compound families, polygamous families, families who were changing occupation, families with dual occupations (agricultural and nonagricultural simultaneously), joint family holdings, and shared cropping (Uganda (1966)).

3. Response Errors

Response errors arise because a respondent is unable to provide the desired information accurately. His social background, level of education, the type of survey, and the reference period play a large part in the occurrence of this type of error.
3.1. Social and educational background

The majority of people in Africa are peasant farmers who are neither literate nor numerate. They operate small holdings to meet their subsistence needs with only a small surplus left for marketing. These farmers do not keep records. They also belong to different language and cultural groups and in some of these groups, it is taboo to count children or cattle (lest they should die). This creates a survey climate in which it is difficult to extract accurate responses, especially numerical ones, as the following examples show.

In many household surveys, responses on such items as size of holdings, farm yield, and incomes have had to be discounted. In Uganda, for instance, it has been reported that peasant farmers have no knowledge of the area they cultivate either in terms of standard units or local units (Mukasa-Mayanja (1976)). In Ghana, it was reported that the results of farmers' interviews have been discounted on the grounds that too few holders had a clear idea of the sizes of their holdings (Kwaku (1986)). In Zambia, the objective estimates for yield rates obtained in the 1970–71 agricultural sample census were for most crops nearly twice as large as the corresponding 1969–70 subjective estimates (except for millet which was four times as large). The reason for this “consistent discrepancy may be either that the farmers who had a very vague idea of the dimension of an acre usually overestimated the size of their fields or there had been a real increase in yield rate from 1969–1970 to 1970–1971” (Zambia (1974)).

In demographic and other household surveys, response errors on age and number of infants and children are common. Age data is highly inaccurate because many Africans do not know their ages or dates of birth, and there is no vital registration to speak of.

Even among those who know their ages, many still misreport it for different reasons. For instance, men tend to exaggerate their ages because age is often venerated. Females, on the other hand, often report themselves to be in the reproductive ages when they are not. These tendencies distort mortality patterns, sex ratios, etc. There has also been a tendency to underreport the number of infants and children for the following reasons.

There is the belief in a number of societies that the “Evil Eye” will affect children if their ages and numbers are exposed to strangers, especially for male children in the age group 5–10. Children are excluded from household counts in societies where infant and early childhood mortality is high. The rationale here is that there is no reason to include someone who will soon pass away.

In some societies, female children are not valued as highly as male children. Female children are underreported or reported as males to confuse the evil spirits (UNECA (1981b)). This tendency has led to the underenumeration of female births.

In food consumption surveys, the method of distributing account books and asking housewives to record the type and quantity of foodstuffs purchased or otherwise obtained does not work because of low literacy rates. Instead, a combination of interviewing and direct measurement of food items is used. Normally, a substantial part of food consumption data is collected through interviews since enumerators are not always present to measure foodstuffs before meals are prepared. To remedy this, housewives are interviewed about the amounts of foodstuffs used to prepare meals. Often this has led to memory errors. The same errors arise in income and expenditure surveys. Households are interviewed about their economic transactions because they do not keep income and expenditure records.

4. Measurement Errors

Measurement errors are especially serious in African agricultural surveys and are discussed in the context of those surveys.
Some African countries still use subjective methods like eye estimation, pacing, etc., to estimate cultivated areas and yields because these methods are less expensive than objective methods. Others use a combination of subjective and objective methods, but the resulting estimates are invariably biased. Most African countries, however, use objective methods and instruments such as chains, tapes, compasses, surveyor arrows, etc., to estimate cultivated areas and yields, at least for principal crops. These methods involve the measurement of randomly selected fields to estimate acreage and crop yield on randomly selected plots of specified shape and size. These methods involve measurement of randomly selected fields in order to estimate acreage and crop cutting on randomly selected plots of specified shape and size in the course of normal harvesting of crops and the subsequent threshing, drying, and weighing for purposes of estimating yield rates. Total crop production is then estimated as a product of cultivated area and yield rate. The application of objective methods has faced a number of problems that in turn lead to measurement errors. It has, for instance, been reported that the “closing errors” of area measurement in the agricultural survey in the Merca district of Somalia were larger than what is recommended by FAO (1981). I discuss some of the causes of measurement errors in agricultural surveys in Africa.

4.1. Area measurement

Inaccuracy in area measurement of fields in African agricultural surveys is caused by many factors. These factors include measuring equipment, shape and size of fields, cropping patterns, and boundary problems, among others.

4.1.1. Inadequate supply of measuring equipment

Shortage of vital equipment has been a problem in a number of countries and this has affected the quality of agricultural surveys. For instance, it has been reported that in Ghana, enumerators have to wait until their colleagues finish using the available equipment. Generally those with equipment hurry through the exercise of measurement so that their colleagues can use the equipment (Kwaku (1986)). But even where equipment is available, the ability of this equipment to record accurate data is sometimes questionable. The measuring wheel, for instance, is of doubtful accuracy when used on wet soil since it tends to get clogged with mud and when used on hilly areas it tends to give too high readings.

4.1.2. Shape and size of fields

In most African countries, fields operated by peasant farmers are small, irregularly shaped, and not cadastrally surveyed. For these reasons these fields are difficult to measure accurately.

4.1.3. Cultivation patterns

The cultivation patterns used in Africa, especially in tropical Africa, tend to create awkward problems both for area and yield measurement.

Mixed cropping, which is the dominant agricultural practice in Africa, involves the planting of several crops at the same time and normally (but not always) harvesting them at the same time. For instance, in Ghana, 84% of the area under seasonal crops contain mixtures of crops. In Botswana, 90% of the area under millet and more than two thirds of the area under sorghum contain other crops (Casley and Lury (1981)). A related practice is associated cropping which involves growing seasonal crops underneath a permanent tree. These agricultural practices make it difficult to give precise meaning to the concept of area. Does one, for instance, take area to be the area of the dominant crop or should one di-
vide the area into half (if there are only two crops)? In Africa, no norms or conventions have been established for apportioning areas where mixed and associated cropping are practiced. There is, however, evidence of enumerators ascribing mixed-crop areas to different crops in an arbitrary manner.

Continuous (or relay) cultivation and continuous harvesting are also common agricultural practices especially in tropical African countries where the rainfall distribution allows for a number of growing seasons per year with considerable overlap between planting and harvesting dates even in a small area (Kiregyera and Dutta Roy (1980)). The holder (farmer) may initially clear a piece of land, plant a crop, and gradually extend his cultivation depending on the availability of family labour. By the time the initial plantings are ready for harvesting, other parts of the field may be at different stages of growth. The cultivation pattern of the field may change several times during the year. In Uganda, for instance, groundnuts are planted from March until May and again from August to October with harvesting in July and November (Casley and Lury, op cit.). The estimates of cultivated areas and yields depend on the time when the area measurements are made. There is also the practice of green harvesting especially of legumes, which starts almost as soon as the green pods appear. The estimates of cultivated areas and yields depend on the time when the area and yield measurements are made. The availability of survey resources normally do not allow for year-round surveillance of crop areas. Consequently, estimates of cultivated areas and yields when there is continuous cropping and harvesting tend to be inaccurate.

4.2. Yield measurement

Crop yield measurement in Africa is complicated and rendered inaccurate by the agricultural practices of mixed cropping, continuous harvesting, shape and size of plots, etc. As already indicated, continuous harvesting makes it difficult to give precise meaning to the concept of yield. A crop is harvested when it is needed. An extreme example of this are the root crops, viz., yams, cassava, potatoes, etc. In good years, these crops remain in the ground, whereas in bad years they are fully harvested before the season has ended. Sampled plots of root crops that are harvested in total can overestimate production. Farmers would not necessarily harvest the entire root crop in a good year (Kiregyera and Dutta Roy (1981)).

The uneven distribution of crops also affects the accuracy of yield measurements. Peasant farmers use the method of "broadcast" rather than planting in rows. This leads to an uneven distribution of crop which usually tends to thin out near the borders. This practice gives rise to a border bias even when the plots are selected at random. It is generally believed that this bias may be negligible if the size of the field is very large compared to the size of the plot. There is, however, inadequate empirical evidence to support this belief.

4.2.1. Shape and size of plots

The shape and size of sample plots for yield measurement can also influence the accuracy of yield estimates. The square seems to be the most used plot shape, although it is known that the square does not have least border bias – the circle does. Research done in India, principally by P.C. Mahalanobis on jute and paddy rice and by P.V. Sukhatme on rice and wheat, indicates that the overestimation is less when plots of moderate sizes are used (Sukhatme (1954)). And the Food and Agriculture Organization of the United Nations (FAO) has studied the sizes of plots for different crops and cultivation patterns. The results of the study show that plot size should be a function of the density of the crop. FAO has
accordingly recommended that, for very dense, irrigated crops, the plot size could be quite small, 1.5 m². For more widely spaced crops (e.g., maize, tubers, etc.) the plot could be as large as 100 m². It would appear that the choice of plot size does not seem to follow the above FAO recommendations and the tendency has been to use smaller plots which tend to overestimate yield. For example, a few common plot sizes would be: 0.92 m² for rice and 1.86 m² for maize in Ghana and 4 m² for maize in Zimbabwe.

5. Interviewer Bias

In Africa and other developing countries no useful information can be collected on a large-scale by self-enumeration. Interviewers are essential and used to ensure that the objectives, concepts and terms, and measurement units are understood by the respondents. Enumerators are also expected to detect inconsistencies and incompleteness of responses, make observations (or measurements) and attempt to minimize refusal and noncompliance. In short, enumerators are considered the keystone of high quality survey work. For instance, research work by Byerlee and Terere (1978) in rural Sierra Leone (West Africa) indicates that the quality of enumerators is important for improving the quality of demographic data. To fulfill their role, interviewers should be conscientious, well-trained, and intensively supervised. They should be able to adhere to the rules for selecting the sample and identifying the sample units. They should be able to carry out measurements according to the stipulated guidelines. And in continuing household surveys, they should be employed on a long term if not permanent basis. Apart from the twenty countries that participate in the African Household Survey Capability Programme (AHSCP) and have a permanent field organization (PFO), ad hoc enumerators are used in most other surveys in Africa. These enumerators are usually students or school drop-outs with no previous experience in data collection. To keep costs low, they will be of junior status and too ill-equipped to handle anything other than simple tools and instructions. Moreover, these ad hoc enumerators are usually not given intensive training or adequate transport (principally bicycles) to canvass their areas. Often supervisors also lack adequate transport to do their jobs efficiently. In the 1986–87 Agricultural Sample Survey in Tanzania, for instance, it was not until two months after the start of the survey that the last bicycles were delivered to the enumerators (Olsson (1987)). There is also abundant evidence that the “morale of these enumerators quickly falls to levels where deliberate under-coverage of fields and taking measurements vitiates the entire survey” (Casley (1986)). These enumerators cannot be expected to do a good job. In addition it has been observed that these enumerators do not follow the instructions closely. For example, in the 1974–75 household budget survey in Zambia, every selected rural household was visited every third day and at each visit, the enumerator was instructed to collect information on: household earnings, expenses, and consumption of own produce. The recall period was the three days that had elapsed since the preceding interview. Some enumerators chose to ask only about the last day’s income and expenditures. Other recorded information on income and expenditure for the previous three days but own produce for just one day, etc. (Zambia (1980)). Measurement of crop production where the measurement requires neither total harvest nor crop cut has created special problems. Many enumerators in agricultural surveys in northern Nigeria simplified their task by weighing only one or two bundles (local units for harvesting sorghum and millet) and fabricating the remaining required data (Poate and Casley (1985)). This practice is clearly a violation of
survey instructions. One of the explanations given for coverage errors in the 1977 Lesotho Fertility Survey is that “interviewers were not sufficiently assiduous in their efforts to contact respondents and that on occasions they may have even deliberately avoided enumerating some households” (Timaeus and Balasubramanian (1984)).

6. Evaluation and Control

6.1. Evaluation

Nonsampling errors arise in household and other surveys despite efforts to control these errors. It is important that these errors are evaluated so that causal factors are identified. This will improve the planning of future surveys and censuses.

There are two ways of evaluating survey data. The indirect method involves comparison of the survey data with reliable data from alternative sources, e.g., a vital registration system. In most African countries it is not possible to find alternative sources of reliable data. As has been pointed out, registration systems are invariably defective and highly unreliable. On the other hand, the direct method involves conducting an ad hoc post-enumeration survey (PES). Only a few PESs have been carried out. These PESs were used to evaluate coverage errors in agricultural and population surveys. For instance, PESs were conducted for the 1965–69 National Sample Census of Agriculture in Malawi, for the 1966 Agricultural Survey in Liberia, and the 1982 Agricultural Sample Census in Botswana. It can be said that inadequate attention has been paid to evaluating household survey data despite that such data is believed to be fraught with large nonsampling errors.

6.2. Control

Control measures to check nonsampling errors have become an essential part of survey design in Africa. These include:

- pretests of survey documents, principally the questionnaire,
- systematic enumeration of dwellings,
- careful recruitment of enumerators and improvement of their training and supervision,
- operational control and reporting system.

Other measures which have been found useful include the creation of a permanent field organization (PFO) with offices in different parts of the country and with a cadre of field staff who are recruited from and reside in these areas. The staff would then know the customs, language, and boundaries of their assigned areas. Other duties that could be assigned to the PFO are: the development of cartographic capabilities, frame construction, updating of population based frames, frame maintenance, acquisition of transport (principally bicycles). Kenya’s National Integrated Sample Survey Programme (NISSP) included a well-documented quality control programme (1974–79). Three types of controls were adopted, namely, coverage, range, and content checks.

The coverage check was designed to find erroneously included units in the frame. The Cartographic Unit and the National Sample Section of the Bureau of Statistics conduct physical checks of cluster boundaries and structure numbering throughout the duration of the NISSP.

The content check was designed to find errors in the data on households in the sample. This work is done by the field supervisors and editing procedures at field level, in district provincial offices and at headquarters. The editing entails identifying those enumerators whose gross-content or gross-coverage error rates were outside the stipulated tolerance limits. Those enumerators whose performance was unacceptable were retrained or asked to re-do the job. In some cases, those on temporary appointment could be dismissed or transferred to an office where they would be under closer supervision.
The range check is performed at the editing stage and checks for outliers during the manual or computer edits. Most of these checks are based on the “reasonableness” of responses on the basis of information given elsewhere.

In addition to the above controls, the minimum educational qualifications for enumerators was raised and their training improved. Also, the role of field supervisors was “re-appraised and redefined with specific emphasis on the need to minimize nonsampling errors” (Kenya (1981)). A low enumerator – supervisor ratio of 4:1 was instituted and supervisors were provided with motorbikes to increase their mobility. Clusters were mapped out and each enumerator was given an “outline map” with which to locate and identify cluster boundaries. Each household was given a unique number and was identified on the map. All dwellings were physically numbered and these numbers entered on listing questionnaires for control purposes. In spite of these control measures it has been reported (Kenya (1981)) that not all surveys under NISSP were successfully implemented. In particular, the following problems were encountered: inaccessibility of clusters, harsh environments, migration of respondents, frequent transfers, inadequate transport, and inadequate equipment. The survey results were withheld from publication when the quality of the data was considered unreliable because of poor enumeration, insufficient training, or poor questionnaire design.

7. References

7.1. References cited in the text


Uganda, Republic of (1966): Report on Ugan-


7.2. References not cited in the text


